DOCUMENT RESUME

ED 385 808 CG 026 396

AUTHOR Beyer, Sylvia; Davanis, Michele
TITLE The Effects of Performance Feedback, Depression, and Gender on Self-Evaluations.
PUB DATE Jun 95
NOTE 21p.; Paper presented at the Annual Meeting of the American Psychological Association (103rd, New York, NY, August 11-15, 1995).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Confidence Testing; Depression (Psychology); Mental Health; Motivation; *Performance Tests; Self Concept; Self Evaluation (Individuals); *Sex Bias; *Sex Differences

ABSTRACT

In an experiment to test whether gender differences in self-evaluation biases exist, 264 female and 174 male students were used. Participants were placed in mixed-sex groups, ranging from 2-10 people. Test anonymity and non-competitiveness were emphasized to minimize self-presentation concerns. Three different measures of accuracy/bias were employed: (1) accuracy of post-test self-evaluations; (2) degree of calibration for individual questions; and (3) response bias. As hypothesized, for the masculine gender-typed test significant gender differences for all three kinds of bias were found: (1) women's post-test self-evaluations were inaccurately low; (2) their confidence statements for individual questions on a test were less well-calibrated than men's; and (3) their response bias was more conservative than men's. None of these gender differences were found for feminine and neutral tests. As hypothesized, strong self-consistency tendencies were found. Expectancies emerged as an important predictor of post-test self-evaluations for both genders and could account for women's inaccurately low self-evaluations. How these biases might negatively affect women's self-confidence and mental health and curtail women's participation in masculine gender-typed domains is discussed. Five graphs and charts displaying gender-differentiated results are appended. Contains 43 references. (SR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 385 808

Running head: SELF-EVALUATION BIASES

The effects of performance feedback, depression, and gender on self-evaluations

Sylvia Beyer

Michele Davanis

University of Wisconsin-Parkside

Poster presented at the annual meeting of the American Psychological Society.
New York City, June 1995

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. BEYER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CG026396

Abstract

This research assessed whether gender differences in self-evaluation biases exist. Three different measures of accuracy/bias were employed: Accuracy of post-test self-evaluations, degree of calibration for individual questions, and response bias. As hypothesized, for the masculine gender-typed test significant gender differences for all three kinds of bias were found: Women's post-test self-evaluations were inaccurately low, their confidence statements for individual questions on a test were less well-calibrated than men's, and their response bias was more conservative than men's. None of these gender differences were found for feminine and neutral tests. As hypothesized, strong self-consistency tendencies were found. Expectancies emerged as an important predictor of post-test self-evaluations for both genders and could account for women's inaccurately low self-evaluations. How these biases might negatively affect women's self-confidence and mental health and curtail women's participation in masculine gender-typed domains is discussed.

The Effect of Gender and Depression on Self-evaluation Biases

Given that a hypothetical male and female have equal abilities, do they perceive their competence similarly or are there gender differences in the accuracy of their self-evaluations of performance? Over the years research on gender differences in self-perception biases has focused extensively on expectancies and causal attributions. It has been suggested that women's low expectancies of performance and their attributional patterns are indicative of their tendency to underestimate their abilities, evidencing a self-derogatory bias (Carr, Thomas, & Mednick, 1985; Erkut, 1983).

The criterion problem

Implicit in the usage of terms such as "underestimation" and "self-derogatory bias" is the assumption that women's expectancies and causal attributions of performance are not only different from men's, but are in fact inaccurate. However, although research has established that gender differences in performance expectancies and causal attributions exist, it has failed to investigate the accuracy of women's and men's self-perceptions. For example, even though women's performance expectancies are lower than men's, their expectancies may be realistic, whereas men's expectancies may be overly optimistic. One possible reason for this neglect of the accuracy question is that objective criteria of accuracy are frequently missing. What constitutes a realistic expectancy of success or an accurate attribution for a performance?

Research on accuracy has a long history within social psychology. Researchers such as Funder (Funder & West, 1993), Kenny (1991), and Kruglanski (1989) have grappled with the problem of how to assess accuracy. These approaches tend to emphasize consensus (interjudge agreement) and self-other agreement as operational definitions of accuracy. Beyer (1990, 1994, 1995) took an alternative path to the assessment of accuracy. She dealt with the criterion problem of self-perceptions by investigating participants' post-test self-evaluations of performance. Accuracy of self-evaluations was operationally defined as the difference between participants' post-test self-evaluations of performance and their actual performance. Beyer found that women's post-test self-evaluations of performance are not only lower than men's but are in fact inaccurately low when compared to their own actual performance, revealing a negative self-perception bias on masculine gender-typed tests of knowledge of sports figures or math. No gender difference in the accuracy of self-evaluations was found for feminine and neutral tests.

Practical implications of accuracy

The ramifications of the accuracy of self-perceptions go far beyond issues of self-knowledge. Entire bodies of literature have investigated the effect of positive illusions about the self on the one hand and depressive realism on the other hand. Positive self-perceptions, even if they are inaccurately high, are related to psychological health (Janoff-Bulman, 1989), whereas negative self-evaluations combined with stressors can lead to depression (Brown, Andrews, Bifulco, & Veiel, 1990). The depressive realism literature suggests that even accurate self-perceptions can have damaging consequences related to depression (Glass, McKnight, & Valdimarsdottir, 1993). Thus, the practical implication of Beyer's (1990, 1994, 1995) findings that women underestimate, i.e., misjudge their own performance on masculine tests, is that self-confidence and psychological health might be adversely affected.

Perceptions of competence are also intimately tied to test performance, persistence, preference for challenging tests, curiosity, intrinsic motivation, expectancies, and aspirations (Boggiano, Main, & Katz, 1988; Cutrona, Cole, Colangelo, Assouline, & Russell, 1994; Grolnick & Slowiaczek, 1994; Harackiewicz & Elliot, 1993; Harter & Connell, 1984). Because perception of competence is an important mediator of achievement behavior, negative self-perception biases are likely to have damaging behavioral consequences which could ultimately become a barrier to success. For example, if one (wrongly) believed that one's past performance was inadequate, why should one seek out a similar test, college course, career, etc. in the future?

Self-consistency and recall biases as explanations of gender differences in self-perceptions

Why might women and men fall prey to different self-perception biases? The theoretical framework used to explain gender differences in self-perceptions in this paper is an extension of self-consistency theory. It is hypothesized that people strive for self-consistency not only in their behavior but also in their judgments of their own abilities. For example, a person's self-conception that s/he lacks

mathematics ability (low performance expectancy for mathematics tests) would negatively affect her/his evaluation of performance on a mathematics test. Expectancies are especially likely to bias self-evaluations of performance when there is ambiguity regarding the quality of performance, such as in the absence of feedback (Wells & Sweeney, 1986). Thus, expectations should have a powerful impact on post-test self-evaluations of performance. Given an identical performance, low expectancies should lead to lower self-evaluations than should high expectancies, especially when no performance feedback is provided. This self-consistency tendency should be attenuated when performance feedback is provided.[1]

Besides self-consistency tendencies, a second process may affect women's self-evaluations negatively. It is possible that when evaluating their performance on masculine tests, women's recall of previously answered questions is biased. Conceivably women remember mostly those questions they believe they answered incorrectly, whereas men remember those questions they believe they answered correctly. This process could bias women towards underestimation of their performance.

The present study

As mentioned above, Beyer (1990) assessed the accuracy of self-evaluations by calculating the discrepancy between performance and self-evaluations, thus employing a reality criterion. The present experiment is designed to add convergent validity to this reality criterion. In addition to conceptually replicating Beyer's (1990) studies by assessing the accuracy of self-evaluations of performance, two other measures of accuracy, namely calibration and response bias, are assessed. If similar gender differences are found using three different measures of accuracy and bias, the results are unlikely to be spurious.

Calibration

In Beyer's (1990) research, participants had to mentally average their performance on many items of a multiple-choice test into one overall post-test self-evaluation of performance. An intriguing question assessed in the present experiment is whether a self-evaluation bias may also operate at the more fundamental level of self-confidence on individual multiple-choice questions. If participants have to state their confidence regarding each question on a test immediately after answering each question, will they be well-calibrated; will they be highly confident when they answered a question correctly but report low confidence when they answered a question incorrectly, rather than vice versa?

Research on calibration has found that individuals are not very well-calibrated. Both sexes tend to overestimate the probability that they answered any particular question correctly (Lichtenstein, Fischhoff, & Phillips, 1982), although a study of adolescents found that girls were better calibrated than boys (Newman, 1984). These results are in contrast to Beyer's findings of women's underestimation of their performance on masculine tests. This discrepancy in findings of accuracy at the level of overall self-evaluations compared to calibration at the level of individual questions may be due to the fact that the tests employed in calibration research tend to consist of general knowledge questions and therefore may be neutral gender-typed. One study that did employ a masculine (math) test found that males are more likely than females to have unrealistically high confidence on math tests (Lundeberg, Fox, & Puncochar, 1994). The present study assessed whether gender differences in the calibration of confidence for individual questions would be found if tests of differing gender-typedness are employed.

Response bias

In addition to the accuracy of self-evaluations of performance and calibration of confidence for individual questions, response bias was assessed using an analysis borrowed from signal detection theory. In the present context, response bias refers to a person's willingness to claim high confidence for having answered a question correctly. Response biases can vary from liberal to conservative, both of which can lead to a high number of errors in self-perceptions. However, the kinds of errors made by persons with different response biases differ. A liberal response bias reveals overconfidence (high confidence about the correctness of an answer to a multiple-choice question, although the question was answered incorrectly). People with a conservative response bias rarely state high confidence regarding the correctness of an

---

[1] The self-consistency hypothesis predicts that expectancies affect self-evaluations above and beyond the effects of performance. However, unless a person is completely out of touch with reality, actual performance imposes boundaries on the self-consistency effect.

answer. They often mistakenly indicate low confidence. If women show inaccurately low post-test self-evaluations of performance on the masculine test in this experiment, it was hypothesized that they would also show a significantly more conservative response bias than men on this test.

Effect of depression

Recent research on depression and self-esteem has recognized the importance of assessing the accuracy of self-perceptions. Some theories of depression (Kovacs & Beck, 1978) and self-esteem (Fitch, 1970) presumed that the self-perceptions of depressives and low self-esteem individuals were negatively distorted. Some research supports this view. Depressed individuals were found to show a bias towards negative information (Bargh & Tota, 1988; Buchwald, 1977; Dykman, Abramson, Alloy, & Hartlage, 1989; Golin & Terrell, 1977; Gotlib, 1983; Johnson, Petzel, Hartney, & Morgan, 1983; Kuiper, 1978; Lobitz & Post, 1979; Roth & Rehm, 1980; Siegel & Alloy, 1990; Wener & Rehm, 1975). However, other research has found that depressives are more accurate than nondepressives in their evaluations of social competence (McNamara & Hackett, 1986), recall of their toddlers' negative behaviors (Lovejoy, 1991), estimates of future success and failure (Alloy & Ahrens, 1987), estimates of positive and negative events that might happen to them (Crocker, Alloy, & Kayne, 1988), and in assessments of the degree of control over external stimuli (Glass, McKnight, & Valdimarsdottir, 1993; Martin, Abramson, & Alloy, 1984). One purpose of this study was to investigate the relation between accuracy and depression.

Hypotheses

In summary, it was hypothesized that gender differences in self-perception biases would be found on measures of accuracy of post-test self-evaluations of performance, calibration, and response bias, but only on a masculine gender-typed test. The gender difference in the accuracy of self-evaluations of performance was hypothesized to be affected by self-consistency and recall biases. The effects of depression and performance feedback on the accuracy of self-evaluations of performance were assessed.

Method

Participants

Participants were 264 female and 170 male students enrolled in introductory psychology courses at the University of Wisconsin-Parkside.

Procedure

Participants were run in mixed-sex groups ranging in size from 2 to 10 participants. To ensure that self-presentation concerns would be minimized, the anonymity of test results and the noncompetitiveness of the tests were emphasized. Participants filled out the Beck Depression Inventory. They then worked on one of three different tests: a masculine gender-typed multiple-choice math test, a feminine English test, and a neutral geography and history test. The tests had been pretested for appropriate gender-typedness. Gender-typedness of test was a between-subjects factor. Each test contained 25 multiple-choice questions. Test presentation and data collection for these tests was accomplished by microcomputer.

Participants were given examples of typical multiple-choice questions for the test they were about to take. They then stated how many questions they expected to answer correctly (expectancies) and rated on a seven-point scale how well they expected to perform, how important it was for them to do well, and how difficult they believed the test to be. Immediately after answering a multiple-choice question, participants had to additionally state how confident they were of having answered that question correctly, henceforth referred to as confidence rating. Participants were familiarized with the confidence rating scale. Confidence ratings could range from "0% sure" indicating that the participant guessed and was completely unsure of the correctness of the answer, to "100% sure" indicating that the participant felt completely sure of having answered the question correctly. After making their confidence ratings for each of the 25 multiple-choice questions, participants estimated the number of correctly answered questions (self-evaluation). They also rated on a seven-point scale how well they believed they had done, how difficult the test was, and stated the number of questions they expected to answer correctly on a similar future test.

Participants then had to recall as many of the multiple-choice questions as possible in 5 minutes and indicated whether they believed they had answered these questions correctly. Finally, they indicated on a seven-point rating scale whether they thought that men or women perform better on each of the three tests.

Results

Analyses of variance (ANOVAs) were performed on the dependent variables. A 2 x 2 x 2 x 3 (gender x feedback condition x depression level x gender-typedness of test) design was employed. For the ANOVAs depression level was operationalized as scores below 10 on the BDI signifying no depression, whereas participants with scores of 10 and above were categorized as depressed (e.g., Anderson, Miller, Riger, Dill, & Sedikides, 1994). The self-consistency hypothesis was tested via multiple regression analyses. Degrees of freedom vary for some of the analyses due to missing values.

Gender differences in the accuracy of self-evaluations

Accuracy of self-evaluations was calculated by subtracting performance from self-evaluation scores. Positive difference scores signify overestimations of performance, negative scores underestimations. ANOVAs were computed with accuracy as dependent variable and gender, feedback condition, depression level and gender-typedness of test as between-subjects factors. The means are shown in Table 1. Level of depression and gender-typedness of test affected the accuracy of self-evaluations, $F(1, 420) = 6.04$, $p < .02$; $F(2, 420) = 35.71$, $p < .0001$. The effect of gender almost reached significance, $F(1, 420) = 3.63$, $p < .06$. The interaction between gender-typedness of test and feedback condition was significant, $F(2, 420) = 18.06$, $p < .0001$. These analyses were followed up with ANOVAs for each test.

Insert Table 1 about here

Women and men did not differ significantly in their accuracy of self-evaluations on the feminine or neutral tests, $F(1, 128) = 1.62$, $p < .21$; $F(1, 119) < 1$, respectively. The gender difference in accuracy of self-evaluations was of marginal significance for the performance feedback masculine test, $F(1, 79) = 3.31$, $p < .08$.

On the feminine and neutral tests, the effect of depression was significant, $F(1, 128) = 4.25$, $p < .05$; $F(1, 119) = 4.13$, $p < .05$. Depressed individuals were more inaccurate (i.e., underestimated) than nondepressed individuals. This finding runs counter the depressive realism hypothesis. On the masculine and neutral tests feedback participants were more accurate than no feedback participants, $F(1, 169) = 21.61$, $p < .0001$; $F(1, 119) = 15.24$, $p < .0001$.

Gender differences in calibration

Calibration was measured using a procedure recommended by Lichtenstein et al. (1982). Confidence ratings between 0% and 20%, 21% and 39%, ... and 80% to 100% were grouped together. Within each of the confidence levels, the percentage of questions that was answered correctly was calculated. For example, assume that participant A indicated "80 to 100% sure" 10 times. However, of those 10 times when A had high confidence, A had answered only 7 questions correctly. In other words, on three questions A was highly confident even though the answer was incorrect. In this example, given a high confidence level, the percentage of questions answered correctly is: (7 /10) x 100 = 70%. To determine statistically whether participants are well-calibrated, a calibration score was calculated. This score indicates "the average deviation of a subject's expressed confidence from the actual proportion of items correctly responded to" (Powel & Nusbaum, 1990, p. 10).[2] Lower scores indicate better calibration. Means can be found in Table 1.

---

[2] The formula for calculating calibration scores is:

$$\text{Calibration} = 1/N \sum_{j=1}^{J} n_j (r_j - c_j)^2$$

where c_j is the proportion of correct responses in confidence category j, r_j is the participant's expressed confidence for answers in category j, n_j is the number of answers in confidence category j, and N is the total number of questions responded to (Powel & Nusbaum, 1990).

ANOVAs were computed with calibration scores as dependent variable and gender, feedback condition, depression level, and gender-typedness of test as between-subjects factors. The main effects for gender, $F(1, 428) = 5.63$, $p < .02$, and test gender-typedness, $F(2, 428) = 3.93$, $p < .02$, were significant. These results were followed up with ANOVAs for each test.

Analogous to the findings for the accuracy of self-evaluations, no significant gender differences in calibration were found on the feminine and neutral tests, $F(1, 129) < 1$; $F(1, 122) < 1$, respectively. As hypothesized, on the masculine test the gender difference in calibration was significant with men being better calibrated than women, $F(1, 171) = 7.23$, $p < .008$. Thus, the absence of findings of significant gender differences in calibration in some of the previous research may have been due to the omission of masculine gender-typed tests from research designs.

Gender differences in response bias

Participants had to discriminate between the correctness and incorrectness of their answers by providing confidence ratings for each multiple-choice question. The question that arises from these data is whether a gender difference in response bias exists. An analysis based on signal detection theory was conducted to answer this question. Signal detection theory is applicable to situations in which there are two discrete states that cannot be easily discriminated (Wickens, 1984). In this study the two discrete states are whether a given question was answered correctly or incorrectly. Participants' confident judgments of having answered a question correctly could fall into two categories. Hits, which represent participants' responses that they are highly confident of the correctness of their answer when they in fact had answered the question correctly and False Alarms, which represent participants' responses that they have high confidence in their answer when in fact their answer was incorrect. Thus, False Alarms represent overconfidence.

In this experiment, response bias, called ß (beta), is the basis upon which participants make the decision to report that they are 80 to 100% certain that they answered a question correctly. Response bias affects the probability of Hits and False Alarms. The more conservative the criterion (the higher ß), the less likely a person is to make high confidence statements. This translates into fewer Hits but also fewer False Alarms. As ß decreases, the criterion becomes more liberal and the person is more likely to claim high confidence. This person will have more Hits but the trade-off is a higher number of False Alarms. Thus, ß is affected by the willingness to claim high confidence when, in fact, the answer was incorrect. By knowing a participant's Hit Rate (HR = Hits / number of correct answers) and False Alarm Rate (FAR = False Alarms / number of wrong answers), one can determine the criterion s/he has set for claiming high confidence.[3]

ß scores are displayed in Table 1. ANOVAs were computed with response bias scores as dependent variables and gender, feedback condition, depression level, and gender-typedness of test as between-subjects factors. The main effect for test gender-typedness was significant, $F(2, 426) = 14.35$, $p < .0001$. There was also a significant test gender-typedness by sex interaction, $F(2, 426) = 4.63$, $p < .01$. These results were followed up with ANOVAs for each test.

As with the other two measures of accuracy, on the feminine and neutral tests no gender difference in ß was found, $F(1, 129) < 1$; $F(1, 122) = 1.77$, $p < .19$, respectively. However, as hypothesized, on the masculine test, the gender difference in ß was significant, $F(1, 171) = 12.79$, $p < .0001$. Females had a significantly more conservative response bias (higher ß) than males. By being less likely to give high confidence ratings, women avoided claiming they answered a question correctly when they had failed to answer it correctly (False Alarm). However, by rarely claiming high confidence, women did not give themselves credit for those questions they did answer correctly (i.e., they had fewer Hits).

To summarize, using three different measures of accuracy and bias, these results provide impressive convergent evidence. On the masculine test women had inaccurately low self-evaluations, were less well-calibrated, and had a more conservative response bias than men. Having demonstrated the

---

[3] A table in Hochhaus (1972) simplifies the mathematical calculations of ß to:

$$ß = ORD(HR) / ORD(FAR)$$

where ORD(p) is the ordinate value of the standardized normal distribution. It should be noted that ß scores are independent of ability (actual performance). Two people with identical performance can have very different ß scores, whereas two people with identical ß scores can have very different levels of performance.

existence of this phenomenon, we will now turn towards an explanation of it.

Self-consistency hypothesis

It was hypothesized that expectancies affect self-evaluations above and beyond the effects of performance (self-consistency effect). This was tested by regressing, for each test, self-evaluation on performance, expectancy, gender, feedback condition, depression scores, and their interaction terms.

Figures 1-3 depict, for each test separately, males' and females' predicted self-evaluations when strategically selected values for performance and expectancies are plugged into the regression equations. Specifically, the average performance score and either a low expectancy score (average expectancy minus one standard deviation), average expectancy score, or high expectancy score (average expectancy plus one standard deviation) were substituted into the regression equations. The line labeled performance indicates the average performance score. If respondents accurately evaluated their performance, their self-evaluations would fall on this straight line. Points below the performance line represent underestimations, whereas points above the performance line represent overestimations.

Insert Figures 1 through 3 about here

As Figures 1-3 clearly indicate, given identical performance, participants' expectancies affect their self-evaluations (witness the steep slopes of the lines). For example, Figure 1 shows that on the feminine test nondepressed, no feedback women with a performance of 18.1 (average performance) have self-evaluations ranging from 14.4 to 21.4 depending on whether they have low, average, or high expectancies. The figures provide visually impressive evidence for self-consistency which is corroborated by the results of the multiple regressions. For the feminine, neutral, and masculine tests expectancy was a highly significant predictor of self-evaluations.

In addition, the graphs illustrate the somewhat lesser effects of gender on self-evaluations (note that the lines for women tend to be below those for men). The graphs also indicate that, as hypothesized, self-evaluations in the feedback condition are more accurate than self-evaluations in the no feedback condition. In addition, depressed individuals have lower and less accurate self-evaluations than do nondepressed individuals.

Overall these findings suggest that performance expectancies are powerful predictors of self-evaluations. However, gender, depression level, and feedback condition are also significant predictors. It should be noted that by knowing a person's performance and expectancy scores, her/his gender, depression score, and whether s/he received performance feedback, we can predict that person's self-evaluations very well. The amount of variance in self-evaluations explained by these variables ranged from 39% to an impressive 92%.

Gender differences in recall

The number of questions a participant recalled as having been answered correctly or incorrectly is expressed as proportion of the participant's total number of recalled questions. It was ascertained whether each recalled question had in fact been answered correctly.

On the masculine test men were more likely than women to recall that a question was answered correctly when it had been answered correctly, $F(1, 170) = 5.57$, $p < .02$, whereas women were more likely than men to believe that they answered a question incorrectly when they had in fact answered it correctly, $F(1, 170) = 7.01$, $p < .009$. On the neutral test, women were more likely than men to believe that they had answered a question incorrectly when they in fact had, $F(1, 123) = 7.44$, $p < .007$. Thus, for women information on failure was more mentally available than for men. This differential recall plus women's reliance on self-consistency may explain women's underestimation of performance on masculine tests.

Discussion

Evidence for gender differences in self-perception biases

The finding that significant gender differences in three different measures of bias (post-test self-evaluations of performance, calibration, and response bias) were found on a masculine test is impressive.

As hypothesized, only on the masculine test did women underestimate their performance more than did men, showed poorer calibration on individual multiple-choice questions than men, and had a significantly more conservative response. This suggests that women have difficulty evaluating their performance on masculine tests. It is important to note that these biases were not found for feminine and neutral tests.

Evidence for self-consistency and biased recall

As hypothesized, expectancies had a significant effect on self-evaluations for all three tests. As Figures 1-3 indicate the self-consistency effect was substantial. This effect was attenuated for participants who received performance feedback. The results suggest that if females have low initial expectancies for mathematics performance, self-consistency tendencies will lead them to an underestimation of their performance in mathematics. In addition, this research and previous research (Beyer, 1995) has shown that when performance is statistically controlled, women still have lower expectancies for future masculine tests. Thus, a self-perpetuating pattern may ensue: Low expectancies lead to inaccurately low self-evaluations which negatively impact expectancies for future test performance which bias self-evaluations, etc. ad infinitum. This process may lead to an avoidance of mathematics (Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1972), which may partially account for the underrepresentation of women in this area (Eccles, 1987). Still, some vexing questions remain. Why do so many females who receive feedback regarding performance in math believe that they did poorly and will continue to do poorly in the future?

The recall data may provide some insight here. On the math test, men were more likely than women to recall questions they believed they answered correctly. Such biased recall is likely to affect self-evaluations. If a relatively low proportion of information on success is mentally available when evaluating one's performance, this should negatively bias self-evaluation. Many of us have known individuals who, after receiving feedback on their performance, focus on and remember the tiny bit of criticism rather than the overwhelming amount of praise. Perhaps females who do well in math focus on the negative aspects of their performance (mistakes) rather than the positive aspects (correct answers), perceive their performance as failure and therefore avoid math in the future.

Warren (1976) also found that women were more likely than men to recall failure viz., anagrams they had not solved. Why might this happen? One may speculate that women focus on the negative aspects of their performance because this information is congruent with their low expectations on masculine tests. In a recent study, Sanbonmatsu, Harpster, Akimoto, and Moulin (1994) found that low self-esteem participants were more affected in their assessments of their verbal intelligence when given negative performance feedback than when given positive performance feedback. High self-esteem participants showed the opposite pattern: Their assessments of their verbal intelligence were more affected by positive rather than negative performance feedback. Assuming that low self-esteem participants have lower expectancies regarding their verbal intelligence (unfortunately this was not assessed), high and low self-esteem participants focused most on expectancy-congruent feedback. Thus, the self-consistency effect may be mediated by memory processes.

Depressive realism hypothesis

Overall, depressed respondents were less accurate self-evaluators than nondepressed respondents: They tended to underestimate their performance.

Because of the serious implications of negative self-perception biases for self-confidence, psychological health, and successful performance, more attention should be devoted to the investigation of gender differences in self-perceptions. A better understanding of the causes of negative self-perception biases may enable us to prevent or at least alleviate these biases which presently may hold back women from achieving their full potential.

References

Alloy, L. B., & Ahrens, A. H. (1987). Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others. Journal of Personality and Social Psychology, 52, 366-378.

Anderson, C. A., Miller, R. S., Riger, A. L., Dill, J. C. & Sedikides, C. (1994). Behavioral and characterological attributional styles as predictors of depression and loneliness: Review, refinement, and test. Journal of Personality and Social Psychology, 66, 549-558.

Bargh, J. A., & Tota, M. E. (1988). Context-dependent automatic processing in depression: Accessibility of negative constructs with regard to self but not others. Journal of Personality and Social Psychology, 54, 925-939.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. Journal of Personality and Social Psychology, 59, 960-970.

Beyer, S. (1994, July). Effects of gender and depression on self-evaluations of performance on academic tasks. Paper presented at the meeting of the American Psychological Society, Washington, D.C.

Beyer, S., & Bowden E. M. (1995). Gender Differences in Self-evaluation Biases. Manuscript submitted for publication.

Boggiano, A. K., Main, D. S., & Katz, P. A. (1988). Children's preference for challenge: The role of perceived competence and control. Journal of Personality and Social Psychology, 54, 134-141.

Brown, G. W., Andrews, B., Bifulco, A., & Veiel, H. (1990). Self-esteem and depression: I. Measurement issues and prediction of onset. Social Psychiatry and Psychiatric Epidemiology, 25, 200-209.

Buchwald, A. M. (1977). Depressive mood and estimates of reinforcement frequency. Journal of Abnormal Psychology, 86, 443-446.

Carr, P. G., Thomas, V. G., & Mednick, M. T. (1985). Evaluation of sex-typed tasks by black men and women. Sex Roles, 13, 311-316.

Crocker, J., Alloy, L. B., & Kayne, N. T. (1988). Attributional style, depression, and perceptions of consensus for events. Journal of Personality and Social Psychology, 54, 840-846.

Cutrona, C. E., Cole, V., Colangelo, N., Assouline, S. G., & Russell, D. W. (1994). Perceived parental social support and academic achievement: An attachment theory perspective. Journal of Personality and Social Psychology, 66, 369-378.

Dykman, B. M., Abramson, L. Y., Alloy, L. B., & Hartlage, S. (1989). Processing of ambiguous and unambiguous feedback by depressed and nondepressed college students: Schematic biases and their implications for depressive realism. Journal of Personality and Social Psychology, 56, 431-445.

Eccles, J. S. (1987). Gender roles and achievement patterns: An expectancy value perspective. In J. M. Reinsich, L. A. Rosenblum, S. A., Sanders (Eds.), Masculinity/femininity: Basic perspectives. Vol 1. New York: Oxford University Press.

Erkut, S. (1983). Exploring sex differences in expectancy, attribution, and academic achievement. Sex Roles, 9, 217-231.

personality judgment: An introduction. Journal of Personality, 61, 457-476.

Glass, D. C., McKnight, D., & Valdimarsdottir, H. (1993). Depression, burnout, and perceptions of control in hospital nurses. Journal of Consulting and Clinical Psychology, 61, 147-155.

Golin, S., & Terrell, F. (1977). Motivational and associative aspects of mild depression in skill and chance tasks. Journal of Abnormal Psychology, 86, 389-401.

Gotlib, I. H. (1983). Perception and recall of interpersonal feedback: Negative bias in depression. Cognitive Therapy and Research, 7, 399-412.

Grolnick, W. S., & Slowiaczek, M. L. (1994). Parents' involvement in children's schooling: A multidimensional conceptualization and motivational model. Child Development, 65, 237-252.

Harackiewicz, J. M., & Elliot, A. J. (1993). Achievement goals and intrinsic motivation. Journal of Personality and Social Psychology, 65, 904-915.

Harter, S., & Connell, J. P. (1984). A model of children's achievement and related self-perceptions of competence, control, and motivational orientation. In J. G. Nicholls (Ed.), Advances in motivation and achievement, Vol. 3 (pp. 219-250). Greenwich, CT: JAI Press.

Hochhaus, L. (1972). A table for the calculation of d' and ß. Psychological Bulletin, 77,

375-376.

Janoff-Bulman, R. (1989). The benefits of illusions, the threat of disillusionment, and the limitations of inaccuracy. Journal of Social and Clinical Psychology, 8, 158-175.

Johnson, J. E., Petzel, T. P., Hartney, L. M., & Morgan, R. A. (1983). Recall of importance ratings of completed and uncompleted tasks as a function of depression. Cognitive Therapy and Research, 7, 51-56.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. Psychological Review, 98, 155-163.

Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. Psychological Bulletin, 106, 395-409.

Kuiper, N. A. (1978). Depression and causal attributions for success and failure. Journal of Personality and Social Psychology, 36, 236-246.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 306-334). New York: Cambridge University Press.

Lobitz, W. C., & Post, R. D. (1979). Parameters of self-reinforcement and depression. Journal of Abnormal Psychology, 88, 33-41.

Lovejoy, M. C. (1991). Maternal depression: Effects on social cognition and behavior in parent-child interactions. Journal of Abnormal Child Psychology, 19, 693-706.

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. Journal of Educational Psychology, 86, 114-121.

Martin, D. J., Abramson, L. Y., & Alloy, L. B. (1984). The illusion of control for self and others in depressed and nondepressed college students. Journal of Personality and Social Psychology, 46, 125-136.

McNamara, K., & Hackett, G. (1986). Gender, sex-type and cognitive distortion: Self-perceptions of social competence among mild depressives. Social Behavior and Personality, 14, 113-121.

Newman, R. S., (1984). Children's numerical skill and judgments of confidence in estimation. Journal of Experimental Child Psychology, 37, 107-123.

Powel, W. D., & Nusbaum, N. (1990, April). Development of accurate error monitoring. Paper presented at the meeting of the Western Psychological Association, Los Angeles, CA.

Roth, D., & Rehm, L. P. (1980). Relationships among self-monitoring processes, memory, and depression. Cognitive Therapy and Research, 4, 149-157.

Sanbonmatsu, D. M., Harpster, L. L., Akimoto, S. A., & Moulin, J. B. (1994). Selectivity in generalizations about self and others from performance. Personality and Social Psychology Bulletin, 20, 358-366.

Siegel, S. J., & Alloy, L. B. (1990). Interpersonal perceptions and consequences of depressive-significant other relationships: A naturalistic study of college roommates. Journal of Abnormal Psychology, 99, 361-373.

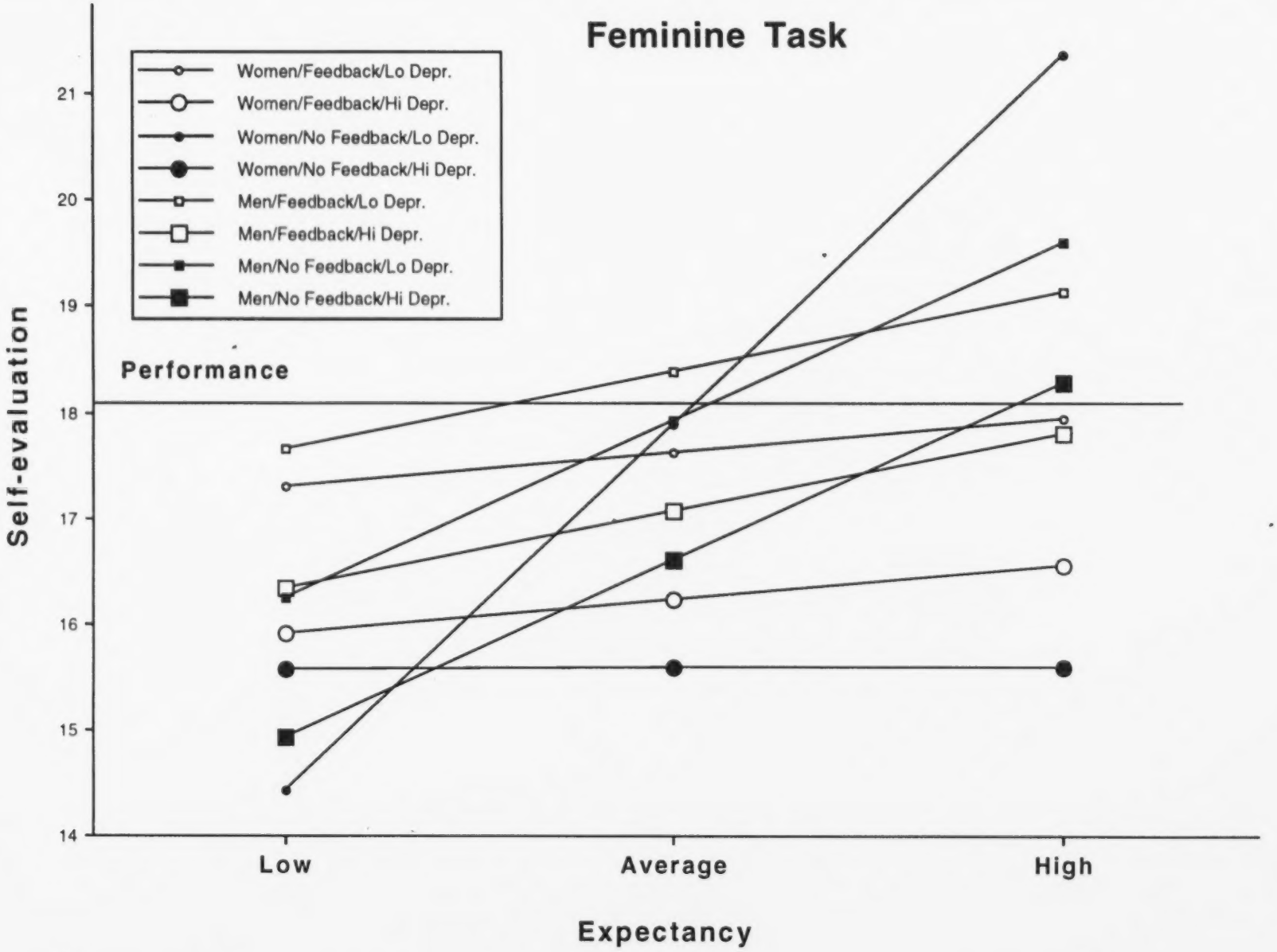
Warren, N. T. (1976). Self-esteem and sources of cognitive bias in the evaluation of past performance. Journal of Consulting and Clinical Psychology, 44, 966-975.

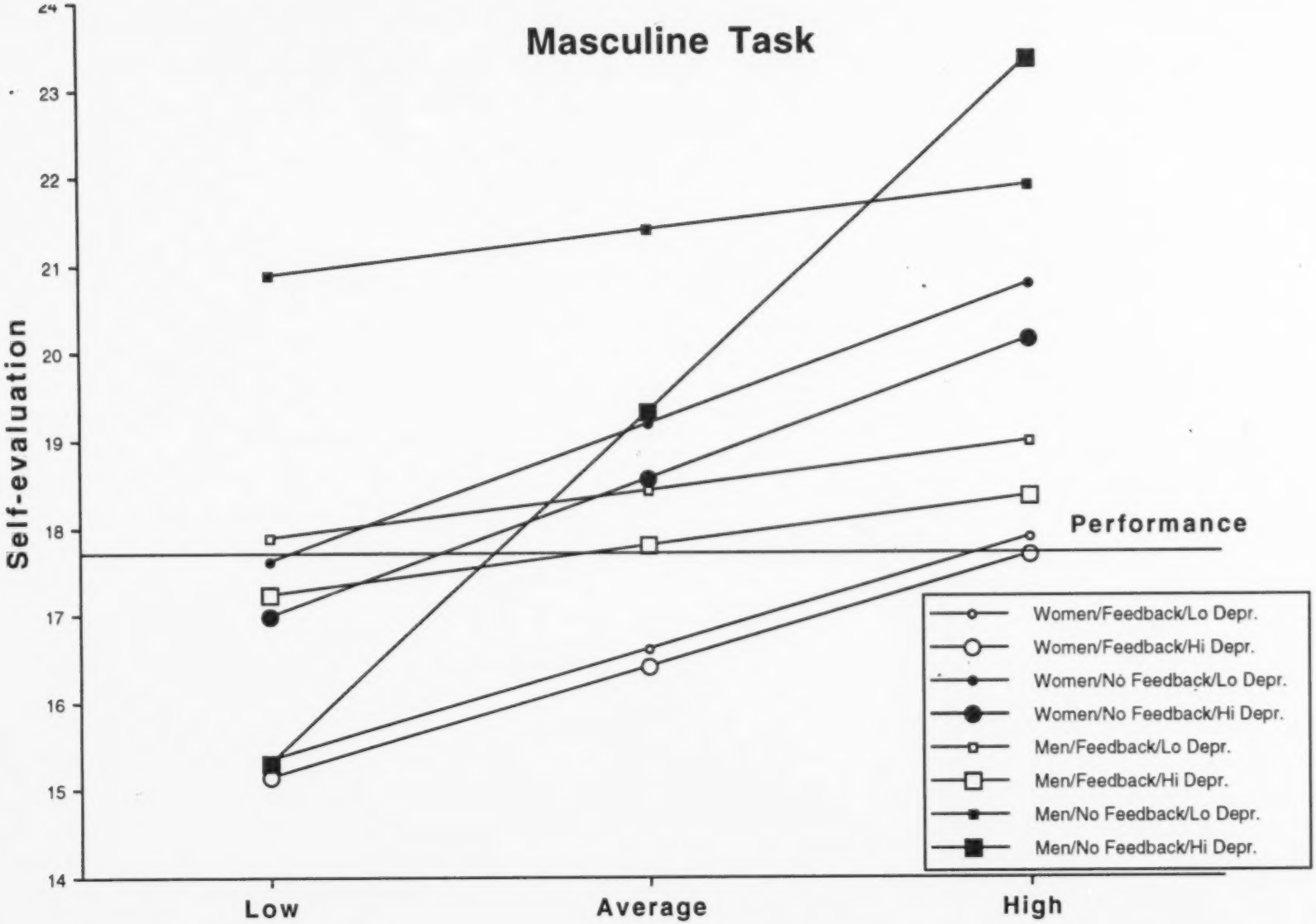
Weiner, B., Frieze, I. H., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1972). Perceiving the causes of success and failure. In E. E. Jones, E. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior. Morristown, NJ: General Learning Press.

Wells, L. E., & Sweeney, P. D. (1986). A test of three models of bias in self-assessment. Social Psychology Quarterly, 49, 1-10.

Wener, A. E., & Rehm, L. P. (1975). Depressive affect: A test of behavioral hypotheses. Journal of Abnormal Psychology, 84, 221-227.

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, OH: Charles E. Merrill.


Feminine Task
Women/Feedback/Lo Depr.
Women/Feedback/Hi Depr.
Women/No Feedback/Lo Depr.
Women/No Feedback/Hi Depr.
Men/Feedback/Lo Depr.
Men/Feedback/Hi Depr.
Men/No Feedback/Lo Depr.
Men/No Feedback/Hi Depr.
Performance
Self-evaluation
21
20
19
18
17
16
15
14
Low
Average
High
Expectancy


Masculine Task
Self-evaluation
24
23
22
21
20
19
18
17
16
15
14
Performance
Women/Feedback/Lo Depr.
Women/Feedback/Hi Depr.
Women/No Feedback/Lo Depr.
Women/No Feedback/Hi Depr.
Men/Feedback/Lo Depr.
Men/Feedback/Hi Depr.
Men/No Feedback/Lo Depr.
Men/No Feedback/Hi Depr.
Low
Average
High
Expectancy

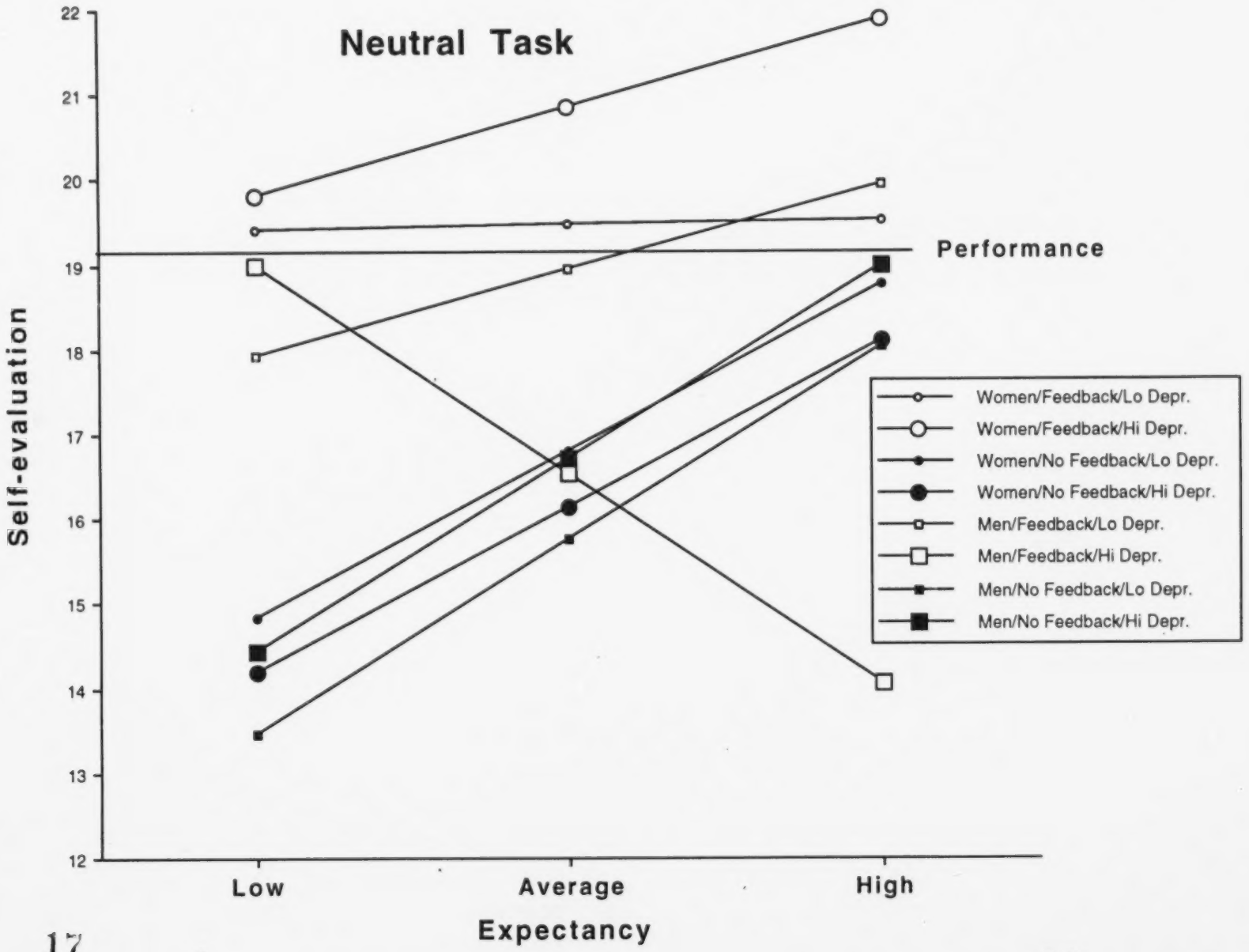

Neutral Task
Self-evaluation
22
21
20
19
18
17
16
15
14
13
12
Performance
Women/Feedback/Lo Depr.
Women/Feedback/Hi Depr.
Women/No Feedback/Lo Depr.
Women/No Feedback/Hi Depr.
Men/Feedback/Lo Depr.
Men/Feedback/Hi Depr.
Men/No Feedback/Lo Depr.
Men/No Feedback/Hi Depr.
Low
Average
High
Expectancy

Table 1. Gender Differences in Various Accuracy Measures

Condition	Accuracy		Calibration		Beta	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
Feminine Task						
No Feedback						
Not Depressed Women (N=26)	-.885	3.983	.038	.029	2.701	6.452
Not Depressed Men (N=20)	.65	2.925	.036	.029	3.603	6.579
Depressed Women (N=15)	-1.467	4.155	.044	.028	4.109	4.186
Depressed Men (N=7)	-2.833	4.262	.045	.027	3.081	2.573
Feedback						
Not Depressed Women (N=27)	-.889	1.761	.05	.029	4.905	4.659
Not Depressed Men (N=21)	-.286	2.004	.046	.032	4.854	5.087
Depressed Women (N=14)	-1.929	2.645	.056	.033	4.282	4.816
Depressed Men (N=3)	-.667	3.512	.028	.016	.099	2.224
Masculine Task						
No Feedback						
Not Depressed Women (N=39)	1.590	3.485	.055	.041	1.768	4.149
Not Depressed Men (N=23)	2.261	3.18	.030	.035	-.671	3.486
Depressed Women (N=23)	1.682	3.695	.054	.052	1.964	3.642
Depressed Men (N=8)	1.000	3.423	.032	.040	-1.731	5.034
Feedback						
Not Depressed Women (N=31)	-.839	2.911	.072	.091	1.554	2.031
Not Depressed Men (N=28)	.148	2.553	.043	.029	.006	5.233
Depressed Women (N=14)	-1.143	3.527	.055	.038	1.454	2.443
Depressed Men (N= 9)	.333	1.581	.06	.035	.51	1.777
Neutral Task						
No Feedback						
Not Depressed Women (N=22)	-2.682	3.483	.056	.041	5.325	6.443
Not Depressed Men (N=20)	-1.895	4.095	.058	.062	6.899	5.117
Depressed Women (N=12)	-2.400	4.115	.067	.021	4.934	4.189
Depressed Men (N=6)	-3.333	4.633	.065	.047	10.581	4.317
Feedback						
Not Depressed Women (N=25)	.320	1.796	.051	.042	6.527	6.876
Not Depressed Men (N=19)	.053	.970	.055	.045	5.864	7.103
Depressed Women (N=16)	-1.938	2.594	.084	.065	5.422	4.029
Depressed Men (N=6)	-.833	1.722	.052	.023	7.850	5.552

Notes. Asterisks indicate significant gender differences * $p < .05$ ** $p < .01$ *** $p < .001$ **** $p < .0001$

Table 2. Gender Differences in Recall of Questions Answered Correctly and Incorrectly

Condition	Recall of							
	Questions Answered Correctly				Questions Answered Incorrectly			
	High Confidence		Low Confidence		High Confidence		Low Confidence	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
Feminine Task								
No Feedback Women	56.8	81.7	10.9	40.7	17.7	46.2	14.5	39.3
Feedback Women	60.1	82.1	3.8	16.1	15.5	41.1	20.7	39.0
No Feedback Men	52.6	84.0	7.5	30.3	21.8	32.2	18.0	35.3
Feedback Men	57.5	88.1	2.8	14.0	17.9	40.5	21.7	43.2
Neutral Task								
No Feedback Women	69.5	100.5	9.2	54.2	9.8	43.1	11.5**	54.5
Feedback Women	73.9	103.2	3.0	26.1	5.6	28.8	17.2**	55.0
No Feedback Men	70.1	76.7	10.2	44.3	11.3	32.4	6.2**	29.4
Feedback Men	78.8	19.6	2.8	10.2	6.2	9.3	11.1**	11.2
Masculine Task								
No Feedback Women	58.1*	90.6	10.3**	36.9	18.0	45.4	13.4	42.5
Feedback Women	58.2*	98.4	9.8**	31.4	11.2	40.0	20.7	39.3
No Feedback Men	72.9*	107.1	4.7**	19.0	15.3	34.5	7.1	22.7
Feedback Men	65.8*	93.1	5.9**	19.4	8.6	28.7	19.8	32.9

Notes. Asterisks indicate significant gender differences * $p < .05$ ** $p < .01$